Genetic profiles of Propionibacterium acnes and identification of a unique transposon with novel insertion sequences in sarcoid and non-sarcoid isolates

AUTHOR(S):

Minegishi, Kana; Watanabe, Takayasu; Furukawa, Asuka; Uchida, Keisuke; Suzuki, Yoshimi; Akashi, Takumi; Maruyama, Fumito; Nakagawa, Ichiro; Eishi, Yoshinobu

CITATION:

Minegishi, Kana ...[et al]. Genetic profiles of Propionibacterium acnes and identification of a unique transposon with novel insertion sequences in sarcoid and non-sarcoid isolates. Scientific Reports 2015, 5: 09832.

ISSUE DATE:

2015-05-12

URL:

http://hdl.handle.net/2433/212469

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
INFECTION
DISEASE GENETICS

Received
2 February 2015

Accepted
12 March 2015

Published
12 May 2015

Correspondence and requests for materials should be addressed to Y.E. (eishi.path@tmd.ac.jp) or F.M. (maruyama.fumito.5e@kyoto-u.ac.jp)

* These authors contributed equally to this work.

Genetic profiles of *Propionibacterium acnes* and identification of a unique transposon with novel insertion sequences in sarcoid and non-sarcoid isolates

Kana Minegishi[1]*, Takayasu Watanabe[2]*, Asuka Furukawa[1], Keisuke Uchida[3], Yoshimi Suzuki[1], Takumi Akashi[4], Fumito Maruyama[5], Ichiro Nakagawa[5] & Yoshinobu Eishi[1,3]

[1]Department of Human Pathology, Tokyo Medical and Dental University Graduate School, 1-5-45 Yushima, Bunkyo-ku, Tokyo, Japan, [2]Department of Periodontology, Tokyo Medical and Dental University Graduate School, 1-5-45 Yushima, Bunkyo-ku, Tokyo, Japan, [3]Department of Pathology, Tokyo Medical and Dental University Hospital, 1-5-45 Yushima, Bunkyo-ku, Tokyo, Japan, [4]Division of Surgical Pathology, Tokyo Medical and Dental University Graduate School, 1-5-45 Yushima, Bunkyo-ku, Tokyo, Japan, [5]Section of Microbiology, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto, Japan.

Propionibacterium acnes **is one of the most commonly implicated etiologic agents of sarcoidosis. We previously reported a complete genome sequence of the C1 strain of *P. acnes* as a clinical isolate from subcutaneous granulomatous inflammatory lesions in a patient with sarcoidosis. In the present study, we initially searched for genetic profiles specific to the C1 strain by core genome analysis and multiple genome alignment with database sequences from 76 and 9 *P. acnes* strains, respectively. The analysis revealed that the C1 strain was phylogenetically independent and carried an 18.8-kbp transposon sequence unique to the sarcoid isolate. The unique composite transposon comprised a novel insertion sequence and extrinsic genes from bacteria other than *P. acnes*. Multilocus sequence typing using 24 sarcoid and 36 non-sarcoid isolates revealed a total of 28 sequence types (STs), including ST26, which was most frequently found without specificity for sarcoid isolates. All 13 ST26 isolates exhibited cell-invasiveness and were confirmed to carry the novel insertion sequence and 4 of the 27 extrinsic CDSs in the transposon, with one exception. ST26 of *P. acnes* with the composite transposon is the most unique strain detected to date and should be further examined as a causative strain of sarcoidosis.**

Propionibacterium acnes is a commensal bacterium on human skin and mucosal surfaces, and is considered causative of acne. Previous studies reported the isolation of *P. acnes* from several tissues, including the conjunctiva, external ear canal, oral cavity, upper respiratory tract, and intestine[1], and the possible association of *P. acnes* with inflammatory disease, such as chronic prostatitis[2], endocarditis[3] and sarcoidosis[4,5,6].

Sarcoidosis is a systemic granulomatous disease with unknown etiology that seems to result from the exposure of a genetically susceptible subject to an environmental agent, and microbial etiologies of sarcoidosis have long been considered based on the clinical similarity to infectious granulomatous diseases[7]. *P. acnes* is the only microorganism isolated from sarcoid lesions by bacterial culture to date[8,9] and one of the most commonly implicated etiologic agents of sarcoidosis[10,11]. A series of Japanese studies proposed an etiology of sarcoidosis as an allergic endogenous infection caused by this indigenous bacterium.

According to the currently-proposed etiology of sarcoidosis[10,11], this low-virulence bacterium causes latent infection in the lungs and lymph nodes and persists in a cell-wall-deficient form. This dormant form of *P. acnes* can be activated endogenously under certain environmental conditions and proliferate in cells at the site of the latent infection. In patients who are hypersensitive to this endogenous bacterium, granulomatous inflammation is triggered by intracellular proliferation of the bacterium. If a certain strain of *P. acnes* causes sarcoidosis, such a causative *P. acnes* strain may have some specific characteristics that confer its intracellular persistency, cell-wall-deficiency, and endogenous activation, or there may be specific antigenicity of the bacterium in sarcoidosis patients.

Many studies of acne vulgaris report that *P. acnes* exhibits phenotypic and genotypic diversity[12,13,14]. In connection with sarcoidosis, Ishige et al. compared genotypes of *P. acnes* strains isolated from the lungs and lymph nodes with those of *P. acnes* indigenous to the skin, conjunctivae, and intestine using random amplified polymorphic DNA analysis[15]. They found that *P. acnes* strains from a particular site were genetically similar, more so than isolates obtained from different sites. Moreover, Minegishi et al. recently determined the complete genome sequence of a *P. acnes* isolate (C1) from granulomatous inflammatory lesions of a patient with cutaneous sarcoidosis[16].

In the present study, we initially performed core genome analysis and multiple genome alignment using the whole genome sequence from the C1 strain of *P. acnes*, compared with 76 and 9 strains of *P. acnes* from a public database, respectively, to search genetic profiles of *P. acnes* from sarcoid tissue samples. In addition, we examined 24 and 36 isolates from sarcoid and non-sarcoid tissue samples, respectively, by multilocus sequence typing (MLST) and polymerase chain reaction (PCR) detection for a *P. acnes*-specific insertion sequence (IS) and extrinsic protein-coding DNA sequences (CDSs) of a novel transposon. The roles of the *P. acnes*-specific transposon with novel ISs and the cell-invasiveness of *P. acnes* with the transposon are discussed in connection with the etiology of sarcoidosis as an allergic endogenous infection caused by this indigenous bacterium.

Results

Monophyly of the C1 sarcoid isolate in core genome analysis. The genomic sequence data for 77 strains of *P. acnes* were available from the database at the time of writing; the C1 strain of *P. acnes* is the only clinical isolate from sarcoid tissue for which the whole genomic sequence has been determined[16]. We first compared amino acid sequences of CDSs among all strains for which genomic sequences were available by sequence similarity.

A total of 1477 single-copy core CDSs were identified, and 1262 of the 1477 CDSs were used for construction of a phylogenetic tree. In the maximum likelihood-based phylogenetic tree, the C1 strain was separately located as a monophyletic clade (Fig. 1), although only the C1 strain was included in the analysis due to unavailability of any other genome information of the isolates from sarcoid tissue samples (sarcoid isolates). These findings suggest that sarcoid isolates have evolved to be monophyletic.

Unique region on the genome of the C1 sarcoid isolate. Next, we compared whole genome sequences of the C1 sarcoid isolate with those from nine other strains available in the NCBI GenBank database using multiple genome alignment (Fig. 2). Homology was observed along whole genome in all dot plots, except inversions in C1 versus ATCC 11828 and C1 versus HL096PA1, which were reported previously[17] (Fig. 2-i). All the breakpoints of these inversions were located in rRNA-encoding regions, and these inversions seemed to occur symmetrically across the replication axes of the genomes. Moreover, the C1 genome contained an 18.8-kbp specific region that was absent in the other nine genomes (Fig. 2-ii). Disruption by this C1-specific region was observed in an alpha/beta hydrolase-encoding CDS, which was intact in the genomes other than C1. At both ends of this region, transposase-encoding CDSs were located between the two similar inverted-repeat sequences; the region was likely a composite transposon composed of two ISs (including transposase between two repeat sequences) at both ends and their intermediate CDSs between the two ISs (Fig. 2-iii).

The ISs of the identified transposon were of a novel family because the sequences were not found in the public database. Max score and e-value of the most similar IS was 44.1 bits and 3e-04, respectively, based on the software ISfinder[18]. The identified transposon contained 13 hypothetical and 14 functionally-known CDSs; the latter included resolvase-coding and arsenate-related CDSs, such as those encoding arsenic resistance protein and arsenite-activated ATPase (Supplementary Table S1).

All 27 CDSs in the transposon were identical to the CDSs of species other than *P. acnes*, such as *P. humerusii*, *P. jensenii*, *P. freudenreichii*, and *P. acidipropionici*.

Phylogenetic dispersiveness of sarcoid and non-sarcoid isolates in MLST analysis. MLST analysis was performed with 24 sarcoid and 36 non-sarcoid isolates, together with the reference ST data (ST1-ST93) available from the public database. The 76 *P. acnes* strains for which either complete or draft genome sequences were available in the public database were excluded in the MLST analysis, because their STs were already known and were therefore less informative (Supplementary Table S2). In a phylogenetic tree constructed from concatenated nucleotide sequences of 9 loci, 28 sequence types (STs) were identified among 60 isolates examined, including novel STs (ST94-ST112) (Table 1 and Fig. 3-i). ST26 isolates were most frequently found in 6 (25%) of 24 sarcoid isolates and 7 (19%) of 36 non-sarcoid isolates without a significant difference between them. The remaining (75%) sarcoid isolates were located dispersively across various STs. The dispersiveness of the sarcoid isolates was supported by differences in the allele number combination shown in the eBURST diagram, although STs of the sarcoid isolates were limited in number (Fig. 3-ii).

ST26 isolates with the novel IS and four CDSs in the transposon. PCR detection of the novel IS was successful in 14 of the 60 isolates including the C1 strain (Table 1). All 13 ST26 isolates and a single ST91 isolate carried the novel IS. Most (12 of 13) of the ST26 isolates carried the four representative CDSs (hypothetical 15.9 kDa protein, arsenic resistance protein, regulatory protein ArsR, and resolvase) that were contained in the unique transposon sequence, with one exceptional strain in which the IS was positive but the four genes were totally negative based on PCR.

ST26 isolates with cell-invasiveness. Comparison of the genomic profiles of *P. acnes* examined in the present study with the cell-invasiveness of each strain reported in our preceding study[19] revealed that 12 (50%) of the sarcoid isolates and 16 (44%) of the non-sarcoid isolates were cell-invasive (Supplementary Table S2). Cell-invasive strains were classified in a limited number of STs (ST8, 26, 36, 41, 67, 70, 100, 112) among a total of 28 STs found in all isolates (Fig. 3-i). All of the ST26 isolates with the novel IS were cell-invasive.

Discussion

We previously reported a whole genome sequence of the C1 strain of *P. acnes* from a granulomatous inflammatory lesion of a sarcoidosis patient[16]. To search for a specific genetic profile of this sarcoid isolate, we first performed core genome analysis with whole genome sequences from 76 *P. acnes* strains and multiple genome alignment with complete genome sequences from 9 *P. acnes* strains. The genomic profiles we found in this sarcoid isolate led to the identification of a transposon unique to the C1 isolate with a novel IS. *P. acnes* strains with the novel IS were classified in ST26 by MLST, with one exception (ST91). PCR analysis for 4 CDSs of the transposon suggested that most of the *P. acnes* strains with the novel IS carry the transposon, which may allow us to determine relevant factors of the bacterium in the etiology of sarcoidosis.

In the present study, ST26 was phylogenetically independent from the other STs based on the core genome analysis (Fig. 1). In the MLST analysis, the ST26 and ST91 strains were phylogenetically independent (Fig. 3-i); however, only ST26, and not ST91, exhibited cell-invasiveness, indicating the phylogenetic independence of ST26 from the others, which was not apparent in the MLST analysis due to the use of only house-keeping CDSs and lack of sufficient genetic information. According to the results by Lomholt and Kilian[12], ST26 of *P. acnes* is

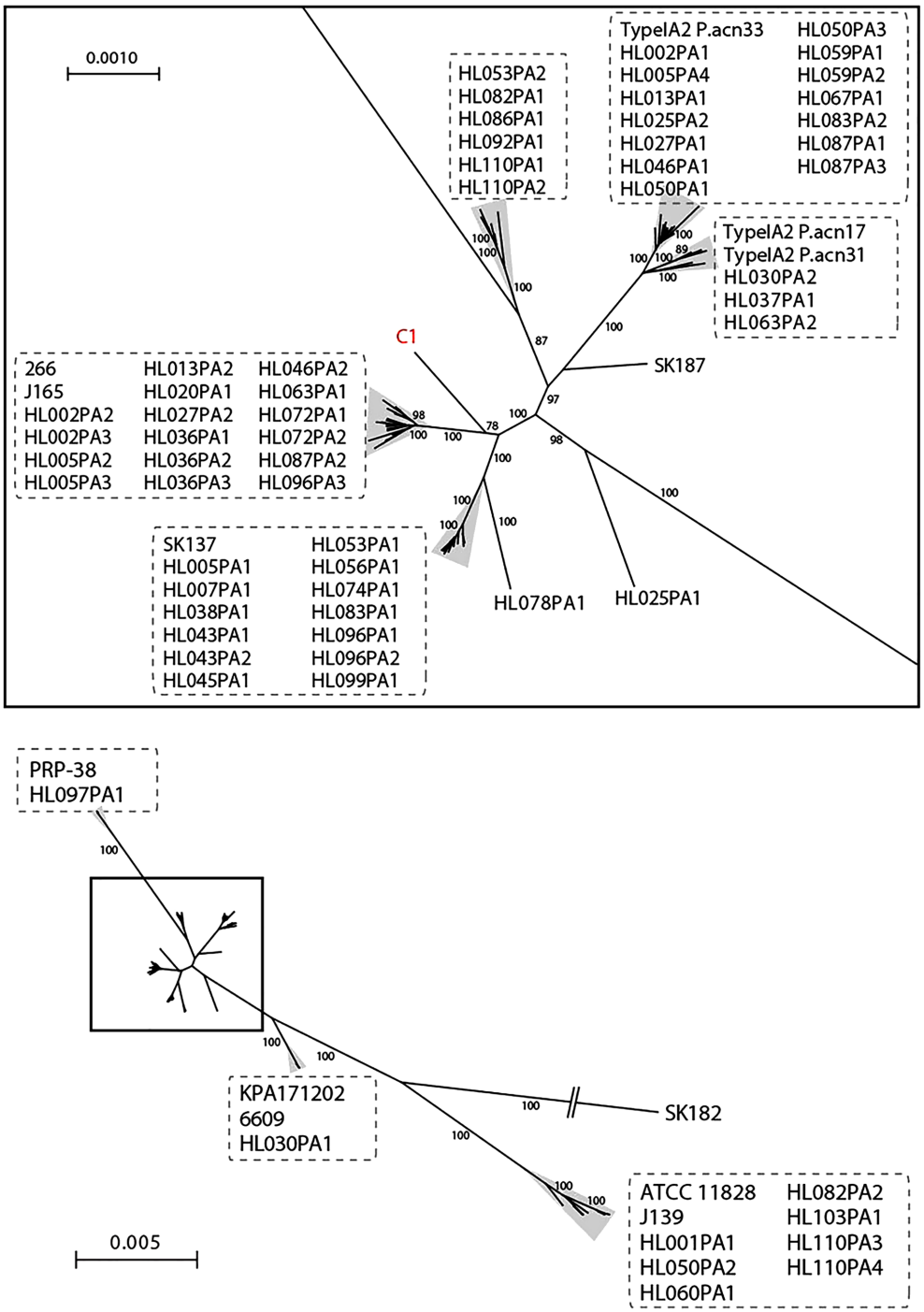

Figure 1 | A maximum likelihood-based phylogenetic tree of 77 *P. acnes* constructed by 1262 core CDSs. The tree was constructed by concatenated amino acid sequences of 1262 core CDSs among 77 *P. acnes* genomes. Detailed tree structure of a dense-branching part is shown in the upper box. Intricate parts in the main and detailed trees are indicated in grey, and the strains in each part are shown altogether without precise location at the tree. Only bootstrap probability values over 70% are given. The sarcoidosis-derived strain is indicated by red color.

different from other ST groups of *P. acnes* in terms of the mutational status of the two hemolytic-associated genes (*camp 5* and *tly*) of this indigenous bacterium. The present study demonstrated the phylogenetic independence of ST26 based on the core genome analysis delineating the *P. acnes* population with high resolution.

A well-known genotype of *P. acnes* is *recA* (types I, II, and III). Each of the *recA* genotypes has a characteristic phenotype and *recA* type I is dominant in isolates from acne vulgaris[20]. Based on the genotype of the *recA* gene against the *P. acnes* isolates in this study, all the ST26 strains in this study were classified as type I. The isolates in type I are prevalent in acne vulgaris and exhibit beta-hemolysis[21]. The study of clustered regularly interspaced short palindromic repeats (CRISPR) in *P. acnes* revealed that CRISPR were present exclusively in types II and III, and differentiated type I from type II[20]. Absence of the CRISPR in type I strains is consistent with the presence of the novel transposon in ST26 strains in type I. The ST26 strains might have evolved to be genetically and phenotypically unique in type I, which is the type possibly evolved from type II.

With regard to the uniqueness of ST26 *P. acnes* strains, it is notable that the presence of the unique transposon in the genome was

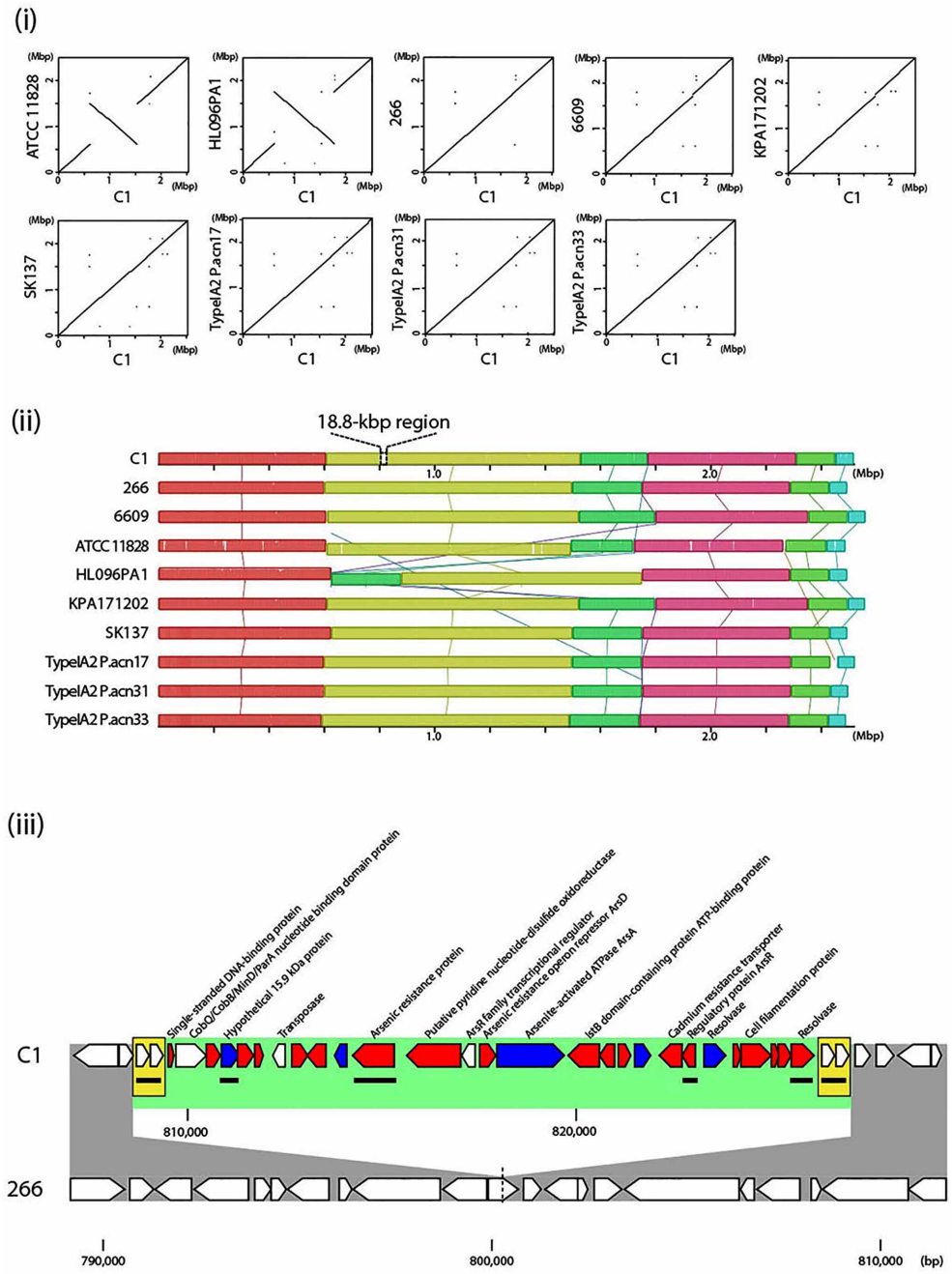

Figure 2 | **Multiple genome alignment of 10 *P. acnes* complete genomes.** (i) Dot plots of C1 against the nine other genomes are shown. Each dot indicates 20-bp match between two genomes, and only ≥ 65-bp continuous dots are shown as a line. (ii) Multiple alignments of 10 *P. acnes* genomes are shown as a figure constructed by Mauve. Each colored box indicates a local collinear block (LCB) that is defined as a genomic region free from genome rearrangements, and LCBs with the same color are linked by lines, indicating homology with each other. An 18.8-kbp C1-specific region is indicated by a dashed line box. (iii) A layout of CDSs in the C1-specific 18.8-kbp region and adjacent region is shown with the corresponding loci on 266 genome. Each boxed arrow indicates a CDS, and the arrowhead is pointed in the transcriptional direction. Homology is indicated in grey, while the novel IS and putative transposon are indicated in yellow and green, respectively. In the transposon, the arrows are colored as follows: red, identical or homologous to the CDS of *Propionibacterium humerusii*; blue, identical or homologous to the CDS of *Propionibacterium sp.* 5U42AFAA strain; white, identical or homologous to the CDS of dairy propionibacteria. Amplicon sites from PCR with the primers in Supplementary Table S3 are indicated by bold lines.

suggested in most ST26 *P. acnes* strains, as well as in the C1 strain. This transposon carried not only functionally known CDSs, such as those encoding arsenical- and metal-resistance proteins, but also hypothetical CDSs for which cell-invasiveness of *P. acnes* seems to be essential for linking this indigenous bacterium to the cause of sarcoidosis, because infectious granulomas are commonly caused by intracellular pathogens. The cell-invasiveness of *P. acnes* is closely associated with the serotype and particular genotypes[19]. In the present study, cell-invasiveness was correlated with a limited number of STs among a total of 28 STs found in the isolates examined. Because all ST26 strains of *P. acnes* were cell-invasive, ST26 strains might have evolved to acquire advantageous characteristics for intracellular persistence of the bacterium after cell-invasion by unknown mechanisms, including horizontal gene transfer via transposition of particular genes. The previous study suggested that specificity of genetic elements to each *P. acnes* lineage contributes to phenotypic and functional differences

Table 1 | Genetic profiles of 60 *P. acnes* strains by MLST and PCR analysis

	Allelic profile											CDSs in the transposon			
Strain name	*cel*	*coa*	*fba*	*gms*	*lac*	*oxc*	*pak*	*recA*	*zno*	ST	Novel IS	Hypothetical 15.9 kDa protein	Arsenic resistance protein	Regulatory protein ArsR	Resolvase
C1	5	9	4	8	4	2	3	5	11	26	+	+	+	+	+
S1	6	11	6	10	6	4	4	9	12	44	−	−	−	−	−
S2	7	11	4	10	1	2	10	6	11	103	−	−	−	−	−
S3	5	4	2	8	4	3	3	5	5	8	−	−	−	−	−
S4	5	4	3	3	4	3	5	1	9	41	−	−	−	−	−
S5	7	11	4	10	1	2	10	6	11	104	−	−	−	−	−
S6	5	9	4	8	4	2	3	5	11	26	+	+	+	+	+
S7	3	9	8	11	7	3	7	6	5	94	−	−	−	−	−
S8	7	9	4	3	4	3	3	2	11	70	−	−	−	−	−
S9	5	9	4	8	4	2	3	5	11	26	+	+	+	+	+
S10	7	9	4	3	4	3	3	2	11	70	−	−	−	−	−
S11	5	4	3	3	4	3	5	1	9	41	−	−	−	−	−
S12	5	4	3	3	4	3	5	1	9	41	−	−	−	−	−
S13	5	9	4	8	4	2	3	5	11	26	+	+	+	+	+
S14	3	13	8	11	7	6	7	6	14	51	−	−	−	−	−
S15	5	9	4	3	4	3	3	2	11	67	−	−	−	−	−
S16	5	9	4	3	4	3	3	2	11	67	−	−	−	−	−
S17	5	9	4	8	4	2	3	5	11	26	+	+	+	+	−
S18	3	13	8	11	7	6	7	6	14	51	−	−	−	−	−
S19	6	11	6	10	6	4	4	9	11	102	−	−	−	−	−
S20	3	13	8	11	7	6	7	6	14	51	−	−	−	−	−
S21	5	9	4	3	4	3	3	2	11	67	−	−	−	−	−
S22	5	9	4	8	4	2	3	5	11	26	+	+	+	+	+
S23	5	4	3	3	4	3	5	1	9	41	−	−	−	−	−
LN1	5	4	3	3	4	3	5	1	9	41	−	−	−	−	−
LN2	3	9	8	11	7	2	5	1	9	95	−	−	−	−	−
LN3	5	9	4	8	4	2	3	5	11	26	+	+	+	+	+
LN4	5	9	4	8	4	2	3	5	11	26	+	+	+	+	+
LN5	5	4	2	8	4	3	3	5	5	8	−	−	−	−	−
LN6	5	9	4	8	4	2	3	5	11	26	+	−	−	−	−
LN7	5	9	4	3	4	3	3	2	11	67	−	−	−	−	−
LN8	3	9	8	11	7	2	7	6	9	105	−	−	−	−	−
LN9	5	9	4	3	4	3	3	2	11	67	−	−	−	−	−
LN10	5	9	4	3	4	3	3	2	11	67	−	−	−	−	−
NS1	5	9	4	8	4	3	3	5	11	100	−	−	−	−	−
NS2	6	11	6	10	6	3	4	5	11	110	−	−	−	−	−
NS3	5	9	4	8	4	3	3	5	11	100	−	−	−	−	−
NS4	7	9	4	3	4	3	3	2	11	70	−	−	−	−	−
NS5	5	9	4	8	4	3	3	5	11	100	−	−	−	−	−
NS6	6	11	6	10	6	4	4	9	11	102	−	−	−	−	−
NS10	5	4	3	3	1	3	3	1	9	98	−	−	−	−	−
NS11	5	9	4	8	4	2	3	5	11	26	+	+	+	+	+
NS12	5	9	4	8	4	2	3	5	11	26	+	+	+	+	+
NS19	5	9	4	8	4	2	3	5	11	26	+	+	+	+	+
NS20	3	13	7	11	7	5	5	6	11	97	−	−	−	−	−
NS21	5	9	4	15	4	2	3	5	11	91	+	+	+	+	+
P1	5	9	4	3	4	3	3	2	11	67	−	−	−	−	−
P2	3	13	7	11	4	3	3	6	14	101	−	−	−	−	−
P3	5	9	3	3	4	3	5	2	9	36	−	−	−	−	−
P4	5	9	3	3	4	3	3	1	9	112	−	−	−	−	−
P6	5	9	1	8	1	3	3	2	11	96	−	−	−	−	−
P8	3	13	8	11	7	3	7	6	14	108	−	−	−	−	−
P9	5	9	4	8	4	2	3	5	11	26	+	+	+	+	−
P10	5	9	2	8	4	3	3	5	5	8	−	−	−	−	−
P11	6	11	6	10	6	3	4	5	11	109	−	−	−	−	−
P12	5	4	3	3	4	3	5	1	9	41	−	−	−	−	−
P14	3	13	8	11	7	3	5	6	13	111	−	−	−	−	−
P15	3	13	7	11	7	3	9	6	11	107	−	−	−	−	−
P16	5	4	3	3	4	3	9	1	9	106	−	−	−	−	−
P17	3	13	7	11	7	3	9	6	14	99	−	−	−	−	−

of *P. acnes* as a commensal and pathogenic agent[22,23]. Considering that a plasmid found in a *P. acnes* strain is suggested to be associated with *P. acnes* virulence[17], the novel transposon might confer novel genetic characteristics to the strains of this unique ST.

The lack of a genetic profile specific to the sarcoid isolates, however, has been reported in several studies. Ishige et al.[15] reported that *P. acnes* isolates were not specific to sarcoidosis when examined by random amplified polymorphic DNA analysis with 45 sarcoid and

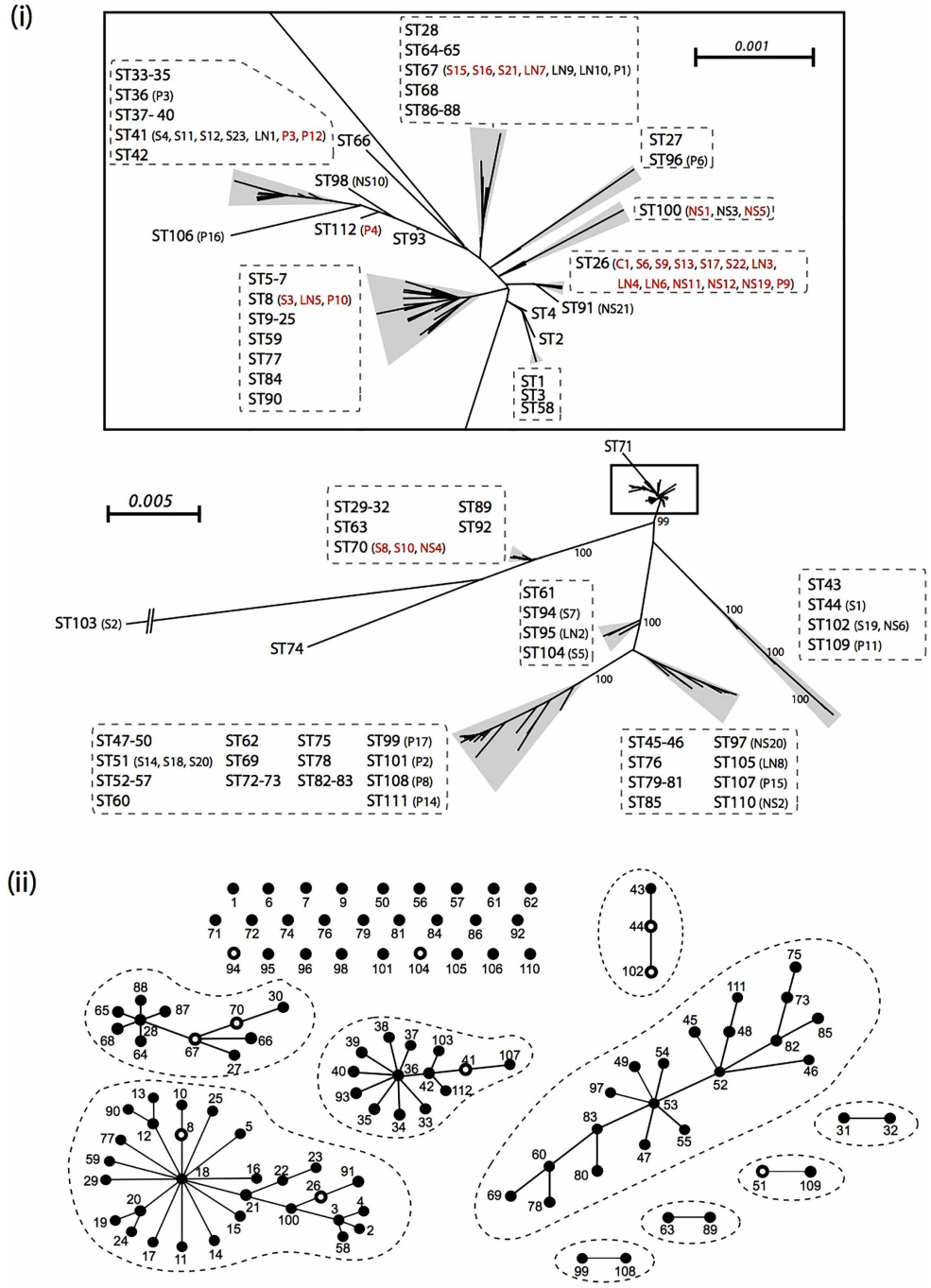

Figure 3 | A neighbor joining-based phylogenetic tree and an allelic profile diagram of 69 *P. acnes* strains and 93 reference STs. (i) The tree was constructed by concatenated nucleotide sequences of nine loci in the *P. acnes* MLST. Detailed tree structure of a dense-branching part is shown in the upper box. The ST types are shown with the isolate names that were classified in the corresponding STs. Intricate parts in the main and detailed trees are indicated in grey, and the strains in each part are shown altogether without precise location in the tree. Only bootstrap probability values over 70% are given. Isolates exhibiting cell-invasiveness are indicated in red. (ii) Diagram constructed by eBURST. Each circle indicates an allelic profile in the *P. acnes* MLST, with the ST number. The ST numbers including sarcoid isolates are indicated by open circles and the others are indicated by filled circles. The STs are single-locus variants against each other if they are linked by a line, and singletons if not linked.

67 non-sarcoid isolates. Furukawa et al.[19] also reported that *P. acnes* isolates were not specific to sarcoidosis in terms of serotype, cell-invasiveness, or genetic polymorphism of the trigger factor gene and the two invasion-associated *P. acnes* genes. Based on the lack of any specific characteristic of the sarcoid isolates, they concluded that host factors that cause an allergic Th1 immune response to the indigenous bacterium are more important for the onset of sarcoidosis than pathogen factors. The present study, however, demonstrated that the sarcoid isolates were likely to have evolved uniquely; the sarcoid isolates might have the capacity to induce chronic inflammation, and unknown factors carried by the transposon unique to the ST26 isolates might be associated with such a characteristic of the sarcoid isolates.

Also, in the present study, ST26 of *P. acnes* with the novel transposon was not specific to sarcoid isolates. The lack of *P. acnes* strains specific to sarcoidosis does not exclude the possibility that a certain strain of *P. acnes* causes sarcoidosis in a genetically susceptible

subject under certain environmental conditions. A single isolate from each sarcoid sample does not always represent the *P. acnes* strain that causes sarcoid lesions due to heterogeneity of characteristics in the population. Most of the sarcoid isolates were cultured from lymph nodes affected by sarcoidosis. This indigenous bacterium is also isolated from some non-sarcoid lymph node samples. Such non-pathogenic strains cannot be discriminated from pathogenic strains when a single colony is picked up from a culture plate as a representative isolate from the sarcoid sample. The C1 strain is an exceptional sarcoid isolate cultured from a sarcoid granulomatous inflammatory lesion in the subcutaneous fatty tissue. Because *P. acnes* has never been found in non-sarcoid subcutaneous tissue, it is free from indigenous flora and seems to be isolated only from the sarcoid granulomatous-inflammatory lesions.

In conclusion, we demonstrated the phylogenetic independency of ST26 strains and their unique characteristics of cell-invasiveness and a unique transposon, and suggested that ST26 is a responsible agent for sarcoidosis. Further studies of ST26 such as whole genome analysis of ST26 *P. acnes* isolates other than C1 are essential for elucidating possible pathogenic factors of this indigenous bacterium in the etiology of sarcoidosis.

Materials and Methods

***P. acnes* strains.** A total of 60 *P. acnes* isolates were evaluated (Supplementary Table S2). All of the *P. acnes* isolates used for the study were collected earlier[15,19]. A representative strain (C1) of sarcoid isolates, which was used for the previous complete genome sequence analysis by Minegishi et al.[16], was isolated from a subcutaneous lesion of a 25-year-old woman with sarcoidosis. Of the other 59 *P. acnes* isolates evaluated, 23 were isolated from 23 lymph nodes of 23 patients with sarcoidosis, 10 were isolated from 10 non-metastatic lymph nodes draining from the stomach, lung, or colon with primary cancer (4, 3, 3 strains, respectively), 12 were isolated from skin swabs of 12 healthy individuals, and 14 were isolated from prostate tissue of 14 patients with prostate cancer. Genomic information of 76 strains (9 complete and 67 draft genomes) was available from the DDBJ/EMBL/GenBank database.

Culture condition and DNA extraction. Stored isolates of *P. acnes* were grown in Gifu anaerobic medium (GAM) broth (Nissui Pharmaceutical Co., Ltd., Tokyo, Japan) at 37°C under anaerobic conditions (10% H_2, 10% CO_2, 80% N_2) for 3 days. Isolation of genomic DNA was described previously[19].

Core genome analysis. All 77 *P. acnes* genome sequences (see above section "*P. acnes* strains"; C1 genome and 76 genomic information in the public database) were processed by the RAST server [24,25] for prediction of CDS regions with functional annotation, and used for the following analysis as the information derived under the same CDS-prediction/annotation criteria. The amino acid sequences of all the predicted CDSs were clustered by PGAP v1.02 under the default parameters[26]. Single-copy core CDSs were identified as those that were located in a single genomic region and commonly present on all the genomes, while strain-specific CDSs were identified as those found exclusively on a single genome. The amino acid sequences of the single-copy core CDSs were concatenated in each strain after exclusion of the CDSs with endogenous rearrangement events by a Phi test, which is a partial algorithm of SpritsTree4[27,28]. The concatenated amino acid sequences were used for construction of a maximum likelihood-based phylogenetic tree. ModelGenerator v851 was used to estimate the appropriate substitution model of amino acid, and RAxML v7.2.8 was used for tree construction under the Jones-Taylor-Thornton model and 100 times bootstrap iteration[29,30]. The tree was visualized by Dendroscope v3[31,32].

Multiple genome alignment. The complete genome sequences of 10 *P. acnes* strains were aligned using the nucmer program in MUMmer v3.23 and progressiveMauve mode in Mauve v2.3.1[33,34]. The C1-specific region was identified as a gap in the Mauve alignment, and ISsaga was used to identify any ISs in the C1-specific region[35]. The identified IS was considered to be novel if all alignments between the identified IS and each of any known ISs in the ISsaga database had <80% length of the known IS and <80% nucleotide identity.

Annotation of the CDSs in the unique transposon. Annotation of the CDSs in the unique transposon was based on the results of BLASTP searches against the NCBI nonredundant protein database[36,37].

PCR conditions. The novel IS and several intermediate CDSs in the transposon were detected by PCR in 59 *P. acnes* isolates using the primers listed in Supplementary Table S3. The PCR conditions for the IS were as follows: 5 min at 94°C, followed by 30 cycles of 30 s at 94°C, 30 s at 58°C, and 80 s at 72°C. The PCR conditions for the other genes were 3 min at 94°C, followed by 30 cycles of 30 s at 94°C, 30 s at 58°C, and 90 s at 72°C. The annealing temperature for arsenic resistance protein was 60°C. The PCR was completed with a final extension step at 72°C for 7 min. Location of the amplicons on C1/266 genomes is shown in Fig. 2.

MLST analyses. We used nine genetic loci (*cel*, *coa*, *fba*, *gms*, *lac*, *oxc*, *pak*, *recA*, and *zno*) for the MLST analyses. PCR conditions and characterization of the allelic profiles were described previously[12,38]. The nucleotide sequences of all nine loci were concatenated to use for the construction of a neighbor joining-based phylogenetic tree by MEGA v5.2 under Kimura's two-parameter (K2P) substitution model and 1000 times bootstrap iteration[39]. The allelic profiles were visualized by drawing a diagram using eBURST v3[40]. The sequence data, and allelic/ST profiles available in the public database (http://pacnes.mlstransposonet) were included in the above MLST analyses.

Nucleotide sequence accession numbers. Nucleotide sequences of the MLST analyses have been deposited in the DDBJ/EMBL/GenBank databases under the following accession numbers: LC006312-LC006851.

1. Funke, G. *et al.* Clinical microbiology of coryneform bacteria. *Clin Microbiol Rev* **10**, 125–159 (1997).
2. Cohen, R. J. *et al. Propionibacterium acnes* associated with inflammation in radical prostatectomy specimens: a possible link to cancer evolution? *J Urol* **173**, 1969–1974 (2005).
3. Vandenbos, F. *et al.* Ventricular patch endocarditis caused by *Propionibacterium acnes*: advantages of gallium scanning. *J Infect* **43**, 249–251 (2001).
4. Ishige, I. *et al.* Quantitative PCR of mycobacterial and propionibacterial DNA in lymph nodes of Japanese patients with sarcoidosis. *Lancet* **354**, 120–123 (1999).
5. Yamada, T. *et al.* In situ localization of *Propionibacterium acnes* DNA in lymph nodes from sarcoidosis patients by signal amplification with catalysed reporter deposition. *J Pathol* **198**, 541–547 (2002).
6. Negi, M. *et al.* Localization of *Propionibacterium acnes* in granulomas supports a possible etiologic link between sarcoidosis and the bacterium. *Mod Pathol* **25**, 1284–1297 (2012).
7. Hunninghake, G. W. *et al.* ATS/ERS/WASOG statement on sarcoidosis. American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis and other Granulomatous Disorders. *Sarcoidosis Vasc Diffuse Lung Dis* **16**, 149–173 (1999).
8. Homma, J. Y. *et al.* Bacteriological investigation on biopsy specimens from patients with sarcoidosis. *Jpn J Exp Med* **48**, 251–255 (1978).
9. Abe, C. *et al.* Frequent isolation of *Propionibacterium acnes* from sarcoidosis lymph nodes. *Zentralbl Bakteriol Mikrobiol Hyg A* **256**, 541–547 (1984).
10. Eishi, Y. Etiologic link between sarcoidosis and *Propionibacterium acnes*. *Respir Investig* **51**, 56–68 (2013).
11. Eishi, Y. Etiologic aspect of sarcoidosis as an allergic endogenous infection caused by *Propionibacterium acnes*. *Biomed Res Int* **935289**, 16 (2013).
12. Lomholt, H. B. & Kilian, M. Population genetic analysis of *Propionibacterium acnes* identifies a subpopulation and epidemic clones associated with acne. *PLoS One* **5**, 0012277 (2010).
13. McDowell, A. *et al.* A novel multilocus sequence typing scheme for the opportunistic pathogen *Propionibacterium acnes* and characterization of type I cell surface-associated antigens. *Microbiology* **157**, 1990–2003 (2011).
14. McDowell, A. *et al.* An expanded multilocus sequence typing scheme for *propionibacterium acnes*: investigation of 'pathogenic', 'commensal' and antibiotic resistant strains. *PLoS One* **7**, 30 (2012).
15. Ishige, I. *et al. Propionibacterium acnes* is the most common bacterium commensal in peripheral lung tissue and mediastinal lymph nodes from subjects without sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* **22**, 33–42 (2005).
16. Minegishi, K. *et al.* Complete genome sequence of a *Propionibacterium acnes* isolate from a sarcoidosis patient. *Genome Announc* **1**, e00016–12 (2013).
17. Kasimatis, G. *et al.* Analysis of complete genomes of *Propionibacterium acnes* reveals a novel plasmid and increased pseudogenes in an acne associated strain. *Biomed Res Int* **918320**, 13 (2013).
18. Siguier, P. *et al.* ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**, D32–36 (2006).
19. Furukawa, A. *et al.* Characterization of *Propionibacterium acnes* isolates from sarcoid and non-sarcoid tissues with special reference to cell-invasiveness, serotype, and trigger factor gene polymorphism. *Microb Pathog* **46**, 80–87 (2009).
20. Bruggemann, H. *et al.* CRISPR/cas Loci of Type II *Propionibacterium acnes* Confer Immunity against Acquisition of Mobile Elements Present in Type I *P. acnes PLoS One* **7**, 3 (2012)
21. McDowell, A. *et al.* A new phylogenetic group of *Propionibacterium acnes*. *J Med Microbiol* **57**, 218–224 (2008)
22. Tomida, S. *et al.* Pan-genome and comparative genome analyses of *propionibacterium acnes* reveal its genomic diversity in the healthy and diseased human skin microbiome. *MBio* **4**, 00003–00013 (2013).
23. Fitz-Gibbon, S. *et al. Propionibacterium acnes* strain populations in the human skin microbiome associated with acne. *J Invest Dermatol* **133**, 2152–2160 (2013).
24. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **42**, 29 (2014).
25. Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 1471–2164 (2008).

26. Zhao, Y. *et al.* PGAP: pan-genomes analysis pipeline. *Bioinformatics* **28**, 416–418 (2012).
27. Bruen, T. C. *et al.* A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681 (2006).
28. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254–267 (2006).
29. Keane, T. M. *et al.* Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol.* **6**, 29 (2006).
30. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014)
31. Huson, D. H. *et al.* Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8**, 460 (2007).
32. Huson, D. H. & Scornavacca, C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* **61**, 1061–1067 (2012).
33. Darling, A. C. *et al.* Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**, 1394–1403 (2004).
34. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12 (2004).
35. Varani, A. *et al.* ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* **12**, 2011–2012 (2011).
36. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402 (1997).
37. Karlin, S. & Altschul, S. F. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc Natl Acad Sci U S A* **87**, 2264–2268 (1990).
38. McDowell, A. *et al. Propionibacterium acnes* types I and II represent phylogenetically distinct groups. *J Clin Microbiol* **43**, 326–334 (2005).
39. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731–2739 (2011).
40. Feil, E. J. *et al.* eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* **186**, 1518–1530 (2004).

Acknowledgments

We thank Ms. Karen Hiratsuka for her technical assistance regarding the MLST analysis. This work was supported by JSPS KAKENHI Grant Number 25670776, 25713060, 24117508.

Author contributions

KM and TW contributed equally to this paper. AF, FM, IN, and YE designed and coordinated the project. KM, KU, YS, and TA performed laboratory experiments. TW, KM, and AF performed the data analysis. KM, TW, FM, and YE wrote the manuscript with assistance from other authors. All authors reviewed the manuscript.

Additional information

Accession codes: The data are available at the DNA Data Bank of Japan (DDBJ) under accession LC006312-LC006851. (http://www.ddbj.nig.ac.jp/).

Supplementary information: accompanies this paper at http://www.nature.com/Scientificreports.

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Minegishi, K. *et al.* Genetic profiles of *Propionibacterium acnes* and identification of a unique transposon with novel insertion sequences in sarcoid and non-sarcoid isolates. *Sci. Rep.* **5**, 9832; DOI:10.1038/srep09832 (2015).